# UECM1703 Introduction to Scientific Computing

## TOPIC 5 Practical

### UNIVERSITI TUNKU ABDUL RAHMAN

| Faculty: | FES | Unit Code: | UECM1703 |
|---|---|---|---|
| Course: | AM &FM | Unit Title: | Introduction Tt Scientific Computing |
| Year: | 1&2 | Lecturer: | Dr Yong Chin Khian |
| Session: | Oct 2022 | | |

Q1.    Consider the following linear system:

$$32.1w + 20.9x + 21.4y + 18.6z = 98.8$$
$$28.8w + 32.4x + 11.6y + 49.4z = 83.4$$
$$69.8w + 49.5x + 44.4y + 15.0z = 138.6$$
$$74.0w + 83.0x + 53.1y + 86.9z = 85.0$$

   (a)    Write the above system in the form $\mathbf{AX} = \mathbf{b}$.

   (b)    Obtain the solution to the system above using matrix inversion.

   (c)    Compute $95.8w + 116.6x + 106.4y + 84.0z$.

Q2.    You are given the following data:

| No. | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|
| 1 | 16.2 | 1.5 | 72.8 |
| 2 | 18.5 | 1.8 | 84.5 |
| 3 | 10.7 | 0.9 | 45.6 |
| 4 | 19.5 | 1.9 | 89.5 |
| 5 | 13.7 | 1.3 | 60.6 |
| 6 | 7.7 | 0.6 | 30.3 |
| 7 | 12.5 | 1.1 | 54.7 |
| 8 | 19.0 | 1.8 | 87.1 |
| 9 | 17.6 | 1.7 | 80.1 |
| 10 | 9.9 | 0.8 | 41.7 |

   (a)    Derive the sample covariance matrix($\mathbf{S_n}$) using the NumPy package.

   (b)    Find the eigen values and eigen vectors of $\mathbf{S_n}$.

   (c)    Find the largest eigen value.

Q3.    A researcher in a scientific foundation wished to evaluate the relation between intermediate and senior level annual salaries of bachelor's and master's level mathematician ($Y$, in thousand dollars) and index of work quality ($X_1$), number of years of experience ($X_2$), and index of publication success ($X_3$). Assume that regression model $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \epsilon$ with independent normal error terms is appropriate.

This practical consists of 4 questions on 4 printed pages

| y | $x_1$ | $x_2$ | $x_3$ | y | $x_1$ | $x_2$ | $x_3$ |
|---|---|---|---|---|---|---|---|
| 33.2 | 3.5 | 9.0 | 6.1 | 40.3 | 5.3 | 20.0 | 6.4 |
| 38.7 | 5.1 | 18.0 | 7.4 | 46.8 | 5.8 | 33.0 | 6.7 |
| 41.4 | 4.2 | 31.0 | 7.5 | 37.5 | 6.0 | 13.0 | 5.9 |
| 39.0 | 6.8 | 25.0 | 6.0 | 40.7 | 5.5 | 30.0 | 4.0 |
| 30.1 | 3.1 | 5.0 | 5.8 | 52.9 | 7.2 | 47.0 | 8.3 |
| 38.2 | 4.5 | 25.0 | 5.0 | 31.8 | 4.9 | 11.0 | 6.4 |
| 43.3 | 8.0 | 23.0 | 7.6 | 44.1 | 6.5 | 35.0 | 7.0 |
| 42.8 | 6.6 | 39.0 | 5.0 | 33.6 | 3.7 | 21.0 | 4.4 |
| 34.2 | 6.2 | 7.0 | 5.5 | 48.0 | 7.0 | 40.0 | 7.0 |
| 38.0 | 4.0 | 35.0 | 6.0 | 35.9 | 4.5 | 23.0 | 3.5 |
| 40.4 | 5.9 | 33.0 | 4.9 | 36.8 | 5.6 | 27.0 | 4.3 |
| 45.2 | 4.8 | 34.0 | 8.0 | 35.1 | 3.9 | 15.0 | 5.0 |

You are given that:

- $n$ is the number of observations.

- $p$ is the number of parameters in the model.

- $\widehat{\boldsymbol{\beta}} = \begin{bmatrix} \hat{\beta}_0 & \hat{\beta}_1 & \hat{\beta}_2 & \hat{\beta}_3 \end{bmatrix}$

- $SSR = \widehat{\boldsymbol{\beta}}^T \mathbf{X^T y} - \frac{1}{n}\mathbf{y^T J y}$, where $\mathbf{J}$ is an $n \times n$ matrix of one.

- $SSE = \mathbf{y^T y} - \widehat{\boldsymbol{\beta}}^T \mathbf{X^T y}$.

- $SST = \mathbf{y^T y} - \frac{1}{n}\mathbf{y^T J y}$.

- $MSE = \frac{SSE}{n-p}$

- $SE(\hat{\beta}_j) = \sqrt{MSE \times C_{jj}}$, where $C_{jj}$ is the diagonal element of the $(\mathbf{X^T X})^{-1}$ corresponding to $\hat{\beta}_j$.

(a) Write the Python commands and output using matrix formulation to obtain the estimate of $\boldsymbol{\beta} = \begin{bmatrix} \beta_0 & \beta_1 & \beta_2 & \beta_3 \end{bmatrix}$.

(b) Determine the predicted value for the mean score of $Y$ with $X_1 = [5.3]$, $X_2 = [20]$, and $X_3 = [6.4]$.

(c) Write the Python commands and outputs to calculated $SSR$.

(d) Write the Python commands and outputs to calculated $SSE$.

(e) Write the Python commands and outputs to calculated $SST$.

(f) You are given that $R^2 = \frac{SSR}{SST}$ and adjusted $R^2$, $R^2_{Adj} = 1 - \frac{SSE/(n-p)}{SST/(n-1)}$. Write the Python commands and outputs to calculated $R^2$ and adjusted $R^2$.

(g) Suppose you are interested to test whether the number of years of experience $(X_2)$ affect salaries, your hypotheses are $H_0 : \beta_2 = 0$ versus $H_1 : \beta_2 > 0$. The corresponding test statistic for testing these hypotheses is $t = \frac{\hat{\beta}_2}{SE(\hat{\beta}_2)}$. Write the Python commands and outputs to calculated $t$.

(h)　Write the Python commands and outputs to calculated the p-value of the test in part(g).

(i)　Write the Python commands and outputs to calculated the test statistic for testing the hypothetses $H_0 : \beta_1 = \beta_2 = \beta = 3 = 0$ versus $H_1 :$ At least one of the $\beta$'s $\neq 0$.

(j)　Write the Python commands and outputs to calculated the p-value of the test in part(i).

Q4.　In an experiment to investigate the effect of color paper (yellow, green and red) on response rates for questionnaires distributed by the "windshield method" in supermarket parking lots, 12 representative supermarket parking lots were chosen in a metropolitan area and each color was assigned random to four of the lots. The reponse rates (in percent) follow.

| | | | $j$ | | |
|---|---|---|---|---|---|
| $i$ | | 1 | 2 | 3 | 4 |
| 1 | Yellow | 28 | 26 | 31 | 33 |
| 2 | Green | 24 | 29 | 25 | 26 |
| 3 | Red | 33 | 38 | 36 | 39 |

Consider the model $y_{ij} = \mu_i + \epsilon_{ij}, i = 1, 2, 3; j = 1, 2, 3, 4$, where

- $y_{ij}$ is the observed response rate for the $i^{th}$ color paper assigned to the $j^{th}$ parking lot.

- $\mu_i$ is the mean response rates of $i^{th}$ color paper.

- $\epsilon_{ij} \sim N(0, \sigma^2)$

Use this model to answer the following questions.

(a)　Let $\boldsymbol{\beta} = (\mu_1, \mu_2, \mu_3)^T$, $\mathbf{y} = [y_{11}, y_{12}, y_{13}, y_{14}, y_{21}, y_{22}, y_{23}, y_{24}, y_{31}, y_{32}, y_{33}, y_{34}]^T$, and $\boldsymbol{\epsilon} = [\epsilon_{11}, \epsilon_{12}, \epsilon_{13}, \epsilon_{14}, \epsilon_{21}, \epsilon_{22}, \epsilon_{23}, \epsilon_{24}, \epsilon_{31}, \epsilon_{32}, \epsilon_{33}, \epsilon_{34}]^T$. The above model can be express in the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Write down the matrix of $\mathbf{X}$ in kronecker form.

(b)　Write down the Python codes and the corresponding output to obtain the estimate of $\boldsymbol{\beta}$, $\mathbf{b}$.

(c)　The hypotheses to test the equalities of means are $H_0 : \mu_1 = \mu_2 = \mu_3$ versus $H_1$: at least one population mean is different from the rest. $H_0$ can be express as $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{d}$. Determine a matrix $\mathbf{C}$ so that $H_0 : \mathbf{C}\boldsymbol{\beta} = \mathbf{0}$.

(d)　Continue using the Python codes from part (b), write down the Python codes and the corresponding output to compute the sum of squares of teaching method, $SST$.

(e)　Continue using the Python codes from part (b), write down the Python codes and the corresponding output to compute the sum of squares of error, $SSE$.

## UECM1703 Introduction to Scientific Computing

    (f)    Continue using the Python codes from part (b), (d) and (e), write down the Python codes and the corresponding output to compute the test statistic.

    (g)    Write down the Python codes and the corresponding output to compute the p-value.